

Appendix

LASSO Methodology Q&A



Below are some detailed responses put together by our team of Brand Licensing Experts for some frequently asked questions.

1. *How have you determined what is “gold-standard?” How many inputs were used in your dataset? How many companies? How did you qualify those companies and products?*

Specifically, each of the experts scored between 28 and 50 brands, totaling 127 brand evaluations of 56 unique brands that served as the “gold-standard” dataset on which the algorithm was trained. These brands belonged to companies in 22 different industries, and included products, services, and media.

2. *The LASSO Model seems highly based on interpretation. Is that right? If so, how do you maintain consistency across scorers, or across your expert panel? Can the LASSO Model be run with true comparative value with any old person scoring the brand?*

It is true that self-scoring based on the LASSO rubric will be subject to personal interpretations of the metric descriptions published in the book and that the scores will be affected by biases common in self-reported surveys. These pitfalls are to some extent unavoidable in this type of self-evaluation, but these drawbacks may be counterbalanced by the ability of

the LASSO scoring assessment to reach a much wider audience and user pool by not requiring a user to retain a brand expert in each case. Regardless, there are mechanisms by which the impact of these response biases and individual interpretations has been blunted.

While it is impossible to phrase any survey question or evaluation description in a perfectly objective manner, guidelines that are specifically defined and neutrally worded will minimize these effects by reducing ambiguity and unintended, unconscious bias. The self-reported LASSO scoring model is unique to many surveys and self-evaluations, in that very detailed descriptions of not just the scoring methods but also the actual basis and background behind the metrics was provided in the chapters of the book. The users, when evaluating their brand, are provided with much more than a few lines describing the scoring guidelines. Rather, they are given a thorough delineation of the concepts that they are being asked to evaluate their brands on. This minimizes the ambiguity that arises when non-experts are required to perform these evaluations and maximizes the consistency in responses. Further, the questions have been phrased in a way that seeks to minimize the emotion involved in evaluating a user's own brand, reducing potential unintended biases on the user's part. Given that any user conducting self-evaluations, and especially non-experts, will always have both biases and individual interpretations regardless of the question's formulation, computational approaches must also be applied to reduce the effects of these confounding factors.¹ By expecting and accounting for these inevitable issues, the LASSO Model is able to predict and model out the effects of these confounding factors to a certain extent. For example, a user evaluating their own

1. Note, this part hasn't been done yet, as we have no data on non-expert user scores. However, this can be easily implemented once the algorithm has been publicly deployed and has over ~20 users.

brand on a quantitative metric for which lower responses indicate a deficiency in their brand or product is highly likely to inflate their score. By having our expert panel assess brands that were also evaluated by non-expert brand owners using the LASSO web application, it is possible to compare responses to the rubric between both the non-biased expert panel and the heavily invested and non-expert brand owners. With enough of these comparisons, a model is able to incorporate this information to predict this overestimation of user scores for self-reported users and subsequently make its final prediction of brand extension more robust to these effects.

3. *How did you make the numbers “relative” across the various sizes of companies and industries they were in? Does this matter here?*

It is certainly important, when training the model, to have a set of brands that represents the diversity of the companies which the users will be trying to evaluate with this algorithm. For example, if only one industry were surveyed, the model would not learn how to use the LASSO metrics to determine a brand’s optimal extension, but rather it would learn how to predict this extension based on arbitrary features of companies in that industry. This would lead to the model performing very poorly in other industries, since these industry-specific features would not be present or useful with these new industries. Note that the training data do not need to contain *every* industry that a user might want to evaluate with the algorithm, but just a diverse enough set so that information from any industry-specific features becomes “drowned-out” relative to the information from the LASSO variables.

Here, brands were selected that the members of the expert panel were familiar with and felt comfortable ranking. The panel members did choose a set of brands that they felt to be representative across companies and industries. Although not perfectly stratified across these domains, over 20 industries were sampled, and no one industry represented more than a

half of the surveyed brands. While there are several industries that were more highly represented than others, most significant being the entertainment industry under which nearly half of the surveyed brands fall, the overall diversity of the training set makes this a reliable dataset to train the algorithm on.

Company size was not specifically controlled for in the training set, and it is a fact that almost all of the brands in this set arise from large companies. This is a consequence of requiring the expert panel members to only consider brands with which they were familiar, in order to ensure that the expert scores were robust and repeatable. This does lead to a potential for the model to perform better on larger companies than smaller companies. Still, given the wide and disparate kinds of companies sampled, across very different industries and product types, the algorithm should be using the information from the LASSO variables to generalize well to companies that it has not seen before. While it is currently trained using brands from these large companies, the model's better predictive ability for larger companies will diminish as the LASSO web application is used more often and the model is able to incorporate information from additional companies across industry, sector, and size.

Industry	Brands (in the Training Set)	Industry	Brands (in the Training Set)
Entertainment	24	Magazine (Home Economics, Interior Design)	1
Consumer Products	10	Electrical	1
Gaming	10	Movies	1
Toys	5	Entertainment (Character)	1

(Continued)

Industry	Brands (in the Training Set)	Industry	Brands (in the Training Set)
Sports	4	Beverage	1
Machinery	3	Sporting Goods	1
Automotive	3	Electronic Manufacturing	1
Nonprofit	2	Fitness	1
Apparel & Fashion	2	Media	1
Restaurant	2	Consumer Electronics	1
Food	2	Food Production	1

Note that brands could be counted twice if they fell within multiple industries.

4. *“...the inclusivity of this training dataset should enable this algorithm to classify accurately brand extension even for industries not present in this dataset.” — This seems like a big claim — almost implausible.*

We understand why this seems to be a grandiose or overconfident statement, but it is rooted in a more formal idea of the ability for a robust model to “generalize” to examples that it hasn’t been trained on yet, even if they are unlike the examples it has been trained on. We kind of touched on this above when talking about how a training dataset does not need to include all industries to generalize well to industries that it hasn’t seen. We’ll try to expand on this a bit here to make this clearer.

Having an inclusive training dataset is important for multiple reasons. The most immediate obvious reason to have as

inclusive of a training dataset as possible is that if an industry is in your training dataset, the model will be trained on it, and the next time the model sees a company or brand from that industry it may be able to apply specific “knowledge” from having been trained on companies in that industry to improve its prediction. However, a less obvious benefit from having an inclusive and representative dataset is that information in the data which is more generally relevant has more of an impact in the model’s training, and the model captures these more widely applicable “ideas” better. If, for example, the model was only trained using brands in the gaming industry, where addictiveness may be overwhelmingly predictive of high brand extensibility regardless of other factors, the model might perform very poorly when faced with brands in the non profit industry where other factors such being Own-able and Storied are also important. However, if the model were trained using examples from both industries, its use of all three metrics in informing its prediction would improve its performance on sports brands, where again all three metrics are highly useful.

A more generic illustration of how using a diverse and inclusive training dataset allows a prediction engine to “generalize” better by considering more relevant features comes from how a young child might learn the definition of a pet. A toddler brought up in a household with only dogs and cats may identify pets as being any animal with four legs, fur, and a tail. A child raised with dogs, cats, and fish, however, would not consider the legs, fur, or tail, but more accurately understand pets to be any animal which the family actively tends to and keeps. Finally, a child who grows up on a farm would correctly learn that pets are animals which the family cares for and takes into their own home, as opposed to animals which are tended to but kept as livestock. In all cases, the “knowledge” learned is not inaccurate, but with more diverse examples the child learns to use features that define the true underlying concept better. When provided with animals that none of the children had seen, the first child may

not identify a caged bird as a pet, while the second child might incorrectly assume that a goat was a pet. Although not guaranteed, the third child would be most likely to categorize both of these examples correctly, despite not having seen them before, due to their learning with more inclusive and diverse “training data.”

5. *“To further improve the accuracy and real-world relevance of the algorithm, a subset of 25 of the brands was independently rated by each of the three experts. This overlap allows the model to capture the intrinsic, yet entirely valid, variation in these metrics. In addition, the overlapping set of examples allows a direct comparison of the agreement between predictions made by the algorithm and those made by human experts.”* — Did a statistician help you create your model?

One of our team members has a significant amount of formal training in statistics at the graduate level, and although he has mostly applied statistics to biological data (he’s a data scientist specializing in genomics), he has a good understanding of the necessary assumptions and best practices behind using these techniques for general data analysis. Regarding the statement here about incorporating information from overlapping training examples from all the experts, it is important to note that for this dataset, we used techniques more commonly classified as machine learning, as opposed to traditional statistics. Although there is considerable overlap between the two fields and a lot of ambiguity over what constitutes their differences, the general difference between the two lies in who selects the features (variables) that are used in the model. In statistics modeling, the data analyst performs this feature selection and manually sets up the model, which is then automatically fitted (trained) using the data. In machine learning, however, both the feature selection and the model training is performed automatically by the computer, with minimal input into the feature selection by the analyst. Both methods have benefits. Because statistical models are

designed by the analyst, it is possible to interpret the model; statistical modeling lets you *explain* the relationship between the variables in the model. However, partly because of their reliance on human curation as well as certain computational limits, they are limited to relatively simple models. Machine learning, on the other hand, strives foremost to *predict* the dependent variable with the best accuracy possible, and because the feature selection and model choice is performed computationally, it is able to generate very complex models that predict complicated phenomena with state-of-the-art results. A consequence of this model complexity, however, is that models generated by advanced machine learning methods usually cannot be interpreted by humans.

We began this analysis trying to use only traditional statistical methods such as logistic regression, because we felt it would be helpful to be able to interpret the effects of the LASSO values on brand extension. However, we quickly found that machine-learning methods performed a lot better for generating predictions in this dataset, as they often do with highly intricate, nonlinear relationships between the variables such as exists here. As an aside, note that we do use our original logistic regression model in the final predictive algorithm, but it is only one “vote” among several other models. We bring all this up because it helps to explain why having this overlap in scores from the expert panel “allows the model to capture the intrinsic, yet entirely valid, variation in these metrics.” If using traditional statistical models, the formal way to add this intrinsic variation in metric scoring between users would be to include an additional random effect feature representing user-judgment in the model. Machine learning, again, does not require or often allow the user to perform this kind of manual feature selection, and simply learns its own features if they improve the final predictions. This is why including these common examples from all three brand experts lets the final model incorporate this additional variation.

6. *“One out of every five cases may be more involved and require further consideration” — How does the LASSO Model help the lay-person distill whether they fall in the 80% camp or the 20% exception camp?*

This is a valid concern, and one that is more difficult to address. Given the vast complexity of determining brand extension and the many intangible factors that affect this phenomenon, it is a challenging problem to objectively quantify and predict. At this point, 80% seems to be the best that can be expected from either human or algorithmic predictors. As more validated training data are collected, the power of big data machine-learning techniques may make it possible to model this phenomenon better and possibly more objectively than even expert humans can; techniques such as artificial neural networks have shown this kind of revolutionary success when given very large, high-quality datasets in many fields, such as business analytics and advertising. In the short term, however, we have a few more techniques to try which may be able to determine if a prediction is correct, even with these fairly small amount of curated data we currently have from this expert panel. Still, there will always be an upper limit to how complex this algorithm can get when trained on small datasets.

7. *“In all, however, the algorithm as it currently exists provides a repeatable, widely-deployable, and inherently objective method for both expert and amateur owners to evaluate their brand.” — How can this be true?*

As discussed above, biases and misinterpretations of the scoring rubric are inevitable in this kind of application, but through both education of the user from the book, well-worded and clear guidelines, and algorithmic correction for biases once data begins to be collected, the effects of these challenges can be minimized. At the end of the day, the availability of this algorithm and online self-evaluation tool will allow much wider adoption of the LASSO Model than would be possible solely through

expert consultation, and the benefits created by this higher accessibility must be weighed against the inaccuracies that go along with it. By observing the mistakes that are common and surveying amateur users of the application, over time it will be possible to incrementally improve the phrasing and user understanding of these metrics alongside the improvements to the algorithm.